DOCUMENT RESUME

ED 292 849 TM 011 229

AUTHOR Schmitt, Alicia P.; Bleistein, Carole A.
TITLE Factors Affecting Differential Item Functioning for Black Examinees on Scholastic Aptitude Test Analogy Items. Research Report No. 87-23.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE Jun 87
NOTE 56p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Black Students; *College Entrance Examinations; Difficulty Level; Factor Analysis; Higher Education; High Schools; *Item Analysis; Standardized Tests; *Test Bias; White Students
IDENTIFIERS Analogies; *Differential Item Performance; *Scholastic Aptitude Test

ABSTRACT

The purpose of this investigation was to identify item factors that may contribute to differential item functioning (DIF) for black examinees on Scholastic Aptitude Test (SAT) analogy items. Initially, items were classified according to several possible explanatory factors. Preliminary analyses identified several factors that seemed to affect DIF for blacks. Secondly, these factors were further investigated by classifying and analyzing analogy items from two additional SAT forms. Results indicate that black students need more time to complete the SAT verbal sections than do their white counterparts with comparable overall SAT verbal scores. Once DIF statistics were corrected for speededness, a smaller number of analogy items were identified as differentially more difficult. Some of the factors appear to be interdependent; these factors include item position within each analogy set, difficulty, subject content, and level of abstractness. The effects of homographs and semantic relationship types are also confounded with the previous factors. Generalizations to be drawn from these results are limited since analyses are based on regular administration items, which tend to have several factors operating in a single item and which contain infrequent occurrence of some factors. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RR-87-23

RESEARCH REPORT

FACTORS AFFECTING DIFFERENTIAL ITEM FUNCTIONING FOR BLACK EXAMINEES ON SCHOLASTIC APTITUDE TEST ANALOGY ITEMS

Alicia P. Schmitt
Carole A. Bleistein

ED292849

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Helen C. Weidenmiller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 011 229

Educational Testing Service
Princeton, New Jersey
June 1987

2. BEST COPY AVAILABLE

Factors Affecting Differential Item Functioning for Black Examinees on Scholastic Aptitude Test Analogy Items[1]

Alicia P. Schmitt
Carole A. Bleistein

Educational Testing Service

[1]This study was supported by Educational Testing Service through Program Research Planning Council Funding.

Copyright © 1987. Educational Testing Service. All rights reserved.

Acknowledgements

The contributions of Neil J. Dorans in specifying the need for this research and his involvement at all stages of the study are gratefully acknowledged.

The authors would also like to recognize the programming and data preparation assistance of Edward Kulick, Ted Blew, and Behroz Maneckshana in completing this study. Acknowledgement for their contributions to the project is also extended to Janice Scheuneman, Edward Curley, and Pamela Chambers for their classification of items and sharing of ideas; to Roger Chaffin, Leslie Pierce and Roy Freedle for permitting the use of their semantic relationship taxonomies and for classifying the items; and to Georgiana Thurston and Karen Damiano for their secretarial assistance. Special appreciation is extended to James Davis, an Educational Testing Service summer graduate student from Cornell University, who examined the items from the perspective of a Black student taking the Scholastic Aptitude Test, and whose insights substantiated the preliminary hypotheses of this study. Additional thanks go to Roy Freedle for sharing the initial data matrices for two of the forms, which permitted a confirmatory phase to this study.

Abstract

The purpose of the present investigation was to identify item factors that may contribute to differential item functioning (DIF) for Black examinees on Scholastic Aptitude Test (SAT) analogy items. This research was considered necessary because analogy items have repeatedly been identified as differentially more difficult for Black examinees. The research was performed in two steps. Initially items in three forms were classified according to several possible explanatory factors. Preliminary analyses identified several factors that seemed to affect DIF for Black examinees. In order to confirm these hypothesized factors, a second step involved classifying and analyzing analogy items from two additional SAT forms.

The most significant finding is that Black students appear to need more time to complete the SAT verbal sections than White students with comparable total SAT verbal scores. This differential speededness effect makes analogy items appear differentially more difficult for Black examinees. Once differential item functioning statistics were corrected for speededness, a smaller number of analogy items were identified as differentially more difficult. In addition, evaluation of the hypothesized factors showed that some of the factors are interdependent and no clear distinction could be made to determine out their individual effects. These item factors are: item position within each analogy set, difficulty, subject matter content, and level of abstractness. The effects of homographs and semantic relationship types are also confounded with the previous factors. The only factor that seemed to be independent was "vertical relationships". In general, a vertical or word associative answering strategy seems to be more consistently used by Black examinees on those items with negative DIF.

Generalizations to be drawn from these results are limited because the analyses are based on regular administration items which tend to have several factors operating in a single item and which contain infrequent occurrence of some factors. At this stage, these results provide a clearer picture of how comparable Black and White students respond on analogy items and the factors that might influence their performance.

Previous research findings have shown that Scholastic Aptitude Test (SAT) analogy items are unexpectedly more difficult for Black examinees than for White examinees with the same level of verbal ability (Dorans, 1982; Kulick, 1984). These results have been replicated across tests and across methodologies for detecting differential item functioning (DIF) but are not readily explained (Echternacht, 1972; Scheuneman, 1978; Stricker, 1982; Rogers & Kulick, 1986).

The purpose of the present investigation was to identify item factors that may contribute to DIF for Black examinees on SAT analogy items. The research was performed in two steps. Initially, the SAT form used in Kulick's (1984) study was examined to generate hypotheses for the DIF found in analogy items. No apparent explanations were observed. Two additional forms were then studied and items were classified according to several possible explanatory factors. Preliminary analyses identified factors which seemed to be related to DIF for Black examinees. In order to confirm these hypothesized factors, a second step involved classifying and analyzing analogy items from two additional SAT forms.

Any source of confusion that may deter examinees from finding analogical relationships could affect performance on analogy items. Preliminary DIF analyses for Black examinees (on three SAT forms) identify the following as possible determinants of DIF on analogy items.

- o differential speededness
- o position within verbal section
- o subject matter content
- o word abstractness
- o relationship between stem and key
- o non-analogical strategies, i.e., vertical relationships or word associations
- o sources of vocabulary confusion, i.e., homographs
- o semantic relationship types

The present study evaluated each of these factors.

Method

Data Source

The data for this investigation come from the SAT-Verbal item responses of White and Black examinees to five SAT forms. Identification of the SAT forms, administration dates and samples by subgroups are presented in Table 1.

Insert Table 1 about here

The verbal test contains 25 Antonym items, 15 Sentence Completion items, 25 Reading Comprehension items, and 20 Analogy items. Items from each type are located in each of the two verbal sections (V1 and V2) which make up the 85-item SAT-Verbal test. Ten analogies are located at the end of V1, the 45-item verbal section, while the other ten are positioned in the middle of V2, the 40-item verbal section.

Performance on the total SAT-Verbal section is reported on the standard College Board scale of 200 to 800. All total verbal sections of the SAT forms used had typical reliabilities (.90's) and met the statistical specifications for difficulty.

Procedure

The standardization method devised by Dorans and Kulick (1986, 1983) was used to study DIF between Black and White test takers. This method identifies unexpected differences in item performance after controlling for ability level differences through standardization. Ability is defined as the total scaled score level on the SAT-Verbal test. The standardization method has been adapted to analyze distractors (including not-reached and omitted responses) and permits variations on how to define the calculation of difficulty values. (See Dorans & Schmitt, 1986.)

Since evaluation of SAT Form OF by Kulick (1984) provided no explanation for DIF in the analogy items, two additional forms (Forms 3H and 4H) were analyzed. Results of these analyses were used to postulate factors that might explain DIF for Black examinees. Two more forms[1] (Forms 3I and 5E) were then examined to validate the hypothesized factors.

Statistics for each item of the 20-item analogy set in each of the five SAT forms used were calculated using the standardization methodology. Distractor analyses were done for four[2] of these five forms. In addition, DIF statistics were calculated by defining difficulty values as total correct divided by all students taking the test ($DSTD_1$) and total correct divided by all students who reached the item ($DSTD_2$).

Results and Discussion

For purposes of this study, results are shown for each SAT form. Forms are ordered by sample ability level, so that results can be interpreted according to the differences of each population (for lower ability examinees harder items will be less discriminating).

Differential Speededness

Distractor analyses demonstrated that analogy items in V1 were differentially reached by Black examinees. Since these items are located at the end of V1 (the longest of the two Verbal sections) which is more speeded than V2, analogy items were more affected. The proportion of Black examinees

---

[1] Initial preparation of the data for these forms was done as part of a study by R. Freedle. Mr. Freedle provided the matrix of the data on which further analyses were performed.

[2] The data matrix from the form used by Kulick (1984) was not available and further analyses could not be performed without incurring additional expense not budgeted for this study.

reaching these items tended to be lower than the proportion of White examinees with the same ability. In order to correct for this effect, the standarization index was recalculated after the proportion correct at each score level was redefined from total correct divided by all students taking the test (P = R/R+W+O+NR), referred to here as $DSTD_1$, to total correct divided by only those students who reached the item (P = R/R+W+O), referred to here as $DSTD_2$, where R = rights, W = wrongs, O = omits, and NR = not reached.

Median values of each of the two standardization DIF indices ($DSTD_1$ and $DSTD_2$) by verbal section and position within section are presented in Table 2.

Insert Table 2 about here

Comparison of DSTD values indicate that if only $DSTD_1$ is considered V1 has a higher negative overall DIF, indicating that Blacks tend to do differentially worse on the V1 analogy items, which appear at the end of V1. After correcting for speededness using $DSTD_2$, this apparent DIF is more in line with that found on the V2 analogy items, which appear in the middle of V2. In addition, $DSTD_1$ and $DSTD_2$ are essentially the same for V2 items, indicating that no correction is needed, i.e., there is no differential speededness effect for V2 analogy items. This is very clear when the two DSTD values are plotted. Figures 1 to 8 show the two DSTD values for each item by Verbal section. The solid line in these figures represents the line of zero difference between same ability Black and White examinees, while the dashed lines indicate $|DSTD| \geq .05$. A $|DSTD| \geq .05$ has been recommended as an useful cutoff for DIF research while a $|DSTD| \geq .10$ identifies problematic items (Dorans & Kulick, 1986).

Insert Figures 1-8 about here

For V1, in all cases, the effect of correcting for differential speededness (using $DSTD_2$) is to make the negative standardized differences less negative and the positive more positive. In other words, negative differential item functioning for Black examinees is not as extreme as that observed prior to the correction.

It is important to note that despite controlling for differential speededness, a number of analogy items remain more difficult than expected for Black examinees. Table 3 summarizes the number of differentially difficult items by Form, item type, DSTD calculation, and DSTD value using DSTD cutoffs of $\leq$ -.05 and $\leq$ -.10.

Insert Table 3 about here

Even after correcting for differential speededness, analogies remain among the two most differentially difficult item types for Black examinees (the other being antonyms), but this difference is not as extreme as it appeared when differential speededness was not taken into consideration. After this correction and considering -.10 as the DSTD cutoff, analogy items appear considerably less problematic for Black students than earlier research indicates. Because of these results, further analyses on the analogy items is reported only for $DSTD_2$.

Position within Verbal Section

The results shown in Table 2 and Figures 1-8 indicate that if the analogy items within the verbal sections are further divided into the first five items and last five items, the first five items account for most of the negative DIF found.

It is important to point out here that analogy items within each 10-item set are organized by difficulty, from easiest to hardest. A trend had previously been observed showing that there is a relationship between item difficulty and differential performance by Black examinees. This possible interaction between differential performance of a subgroup and item difficulty has been reported by Shepard, Camilli, and Williams (1984) and by Rogers and Kulick (1986) in studies of Black differences and by Schmitt (in press) in a study of Hispanic differences.

Specifically, when using the standardization method, part of this apparent relationship between item difficulty and DIF is a function of the weights used in the differential item functioning analyses, which are the relative frequencies of the Black sub-population at each scaled score level of the SAT verbal test. Since Black examinees are a lower scoring group, p-differences between comparable Black and White examinees on easy items will have a greater weight than on hard items. Standardization places the focus on the items which are most relevant for the group. Also, hard items have little impact on group performance, contributing very little to the difference between group means. Other factors might also contribute to making eaiser items appear more difficult for Black examinees. Some of these factors are discussed subsequently.

Subject Matter Content

Test development content specifications were used to analyze the effect of subject matter content on DIF. The four subject matter content classifications used are:

- o Aesthetics or Philosophy - which includes art, literature, drama, music, religion and other related subjects.

- o World of practical affairs - which includes economics, politics, sports and other related subjects.
- o Science - which includes biology, applied science, agriculture, medicine and other related subjects.
- o Human relationships - such as emotions, family, character analysis, psychology, etc.

Differential item function statistics for analogy items in each verbal section and total test for the four subject matter content classifications are presented in Table 4.

Insert Table 4 about here

Results in Table 4 show that the science subject matter content classification has the highest negative DIF of all classifications and that this negative DIF is consistent across forms. Inspection of the positions where science content items are located indicates that in all forms a higher proportion are among the first five items of each 10-item set (3I-4/1, 3H-3/2, 4H-4/2, 5E-4/1) where the fraction after each form code represents frequencies by position (numerator - first five positions; denominator - last five positions).

Human relationship subject-matter content shows the highest and most consistent positive DIF, indicating that Black students tend to do better on items whose content is about human relationships. In this classification, as with science, there seems to be an interaction with position. More human relationship items are positioned as the last five items of each set: (3I-1/3, 3H-2/3, 4H-2/2, 5E-0/5.)

The interrelationship among subject content, position, and difficulty could explain why items exhibit differential functioning. For example, more abstract

words are used in the hardest items, which generally appear among the last five items of each analogy set. In addition, abstract terms are used to express human relationships. On the other hand, science related words tend to be concrete and be used on easier items, which are located among the first five items of each analogy set.

Word of Abstractness

As implied by the previous explanations, the abstractness of the words in the items might also have an effect on DIF. Test development classifications were used to evaluate word abstractness. The categories used are defined as:

- o Concrete - all words in the stem and key represent concrete entities.
- o Mixed - words in the stem and key are a combination of concrete and abstract entities.
- o Abstract - all words in the stem and key represent intangible entities.

Table 5 presents median values of DIF statistics for analogy items by verbal section for the abstract/concrete classification.

Insert Table 5 about here

Results show that items with concrete words tend to have more extreme negative DIF. Inspection of where these items are located indicates that they tend to be most consistently positioned as the first five items of each set. (3I-6/0, 3H-5/0, 4H-4/1, 5E-5/0).

Abstract words are usually used in analogy items which show consistent positive DIF by Black examinees. These items tend to be located in the last five positions of each 10-item analogy set. (3I-1/6, 3H-1/7, 4H-3/3, 5E-0/7).

As with subject matter content, word abstractness has an interdependence with position and item difficulty as factors influencing DIF. This interdependence prevents a clear discrimination of the effect of each individual factor on DIF.

Relationship between Stem and Key

A third classification by test development categorizes items according to whether or not the stem and key overlap. The two categories are:

- o Independent - neither word in the key is suggestive of a word in stem.
- o Overlapping - one or both of the words in the key is suggestive of one or both words in the stem.

Results of DIF for analogy items categorized into "independent" or "overlapping" by forms and verbal sections are presented in Table 6.

Insert Table 6 about here

Results for the "independent" or "overlapping" categories are not consistent and seem to be mostly dependent on item position. For items classified as "independent" the proportion of first position to last position is: 3I-5/8; 3H-7/9; 4H-5/8; 5E-5/7. As seen by the DIF medians on Table 5 for the "independent" category there is almost no DIF. For the "overlapping" category there is slightly more negative DIF but it interacts with position. (3I-5/2; 3H-3/1; 4H-5/2; 5E-5/3.)

Vertical Relationships

Another way to categorize the relationship between the words of the stem, key and distractors can be defined as a vertical relationship. Vertical-word associations are non-analogical answering strategies where the relationship between the terms on the analogy item stem are not considered when selecting an

option as correct. Typically, an analogy item is composed of S:T :: O:P (where S:T is the stem and O:P is any of the five options). A vertical relationship will be made when an association is found between S and any O or T and any P. For example:

PHYSICIAN:PATIENT:: (A) nurse:hospital
(B) guard:warden (C) informer:agent
(D) attorney:lawyer (E) accountant:client

The desired relationship in this item is one of subject to recipient, but the non-analogical vertical type of relationship is a medical one, making distractor A (nurse:hospital) an attractive distractor. Use of such incorrect non-analogical answering strategies which involve vertical association between words could be used more by examinees who are less experienced with analogies.

Each word of the key and distractors for each analogy item was classified in terms of its one-to-one relationship to the stem. For example, in the analogy-item stem "SOCCER:TEAM" the term "team" in the stem would have a vertical relationship to "coach" and to "players" in a distractor, in that coach and players are part of a team. This distractor would receive a code of 2 for the two vertical relationships. In this way, for each item, its key and each of its distractors could be classified as having from zero to two vertical relationships. In addition, for purposes of analysis, a total vertical relationship was obtained for the distractors by adding all distractor codes. Each word of the 80 analogy items was classified by at least four people: by two test development specialists and by the two primary investigators. Consensus of all four people was reached for any discrepancies. Results of $DSTD_2$ medians for vertical relationships in the key and in all distractors of analogy items by form and verbal section is presented in Table 7.

Insert Table 7 about here

Due to the small number of vertical relationships in some categories, the codes were collapsed to present results of DIF medians. Vertical relationships for the key were defined as occurring or not. If at least one word in the key was vertically related to the stem, the key was given a code of one; if no word in the key was vertically related to the stem, the key was given a code of zero.

It was expected that, if vertical relationships were used more by Black examinees when the relationship was located in the key, then Black candidates should perform differentially better than White examinees on those items. Results for vertical relationships on the key show that the medians for either code are not consistent. Examination of item position helps interpret results further. The proportion of items in each position for those items where there is a vertical relationship in the key is: 3I-6/3, 3H-4/2, 4H-5/2, and 5E-4/2. More than twice the number of items where there is a vertical relationship in the key are located in the first five positions of each verbal sections. Only for Form 3H is the median $DSTD_2$ values somewhat negative (-.02).

Categories of vertical relationships for distractors were added across all distractors and further collapsed for the category represented by a two. This category includes two and three vertical relationships. Category three represents those items where more than three vertical relationships were identified across its distractors. Results indicate that the expected negative DIF for more vertical relationships in the distractors is observed. It is possible that distractors might become more attractive because of the vertical relationships preventing selection of the key or use of the real analogical relationship. As the number of vertical relationships increase, the DIF medians are more negative.

Homographs

Words which are spelled alike but have different meanings in English, i.e., homographs, have been identified as a factor that might make items unexpectedly more difficult for some minority groups such as Asian-Americans (Bleistein & Wright, 1986) and Hispanics (Schmitt, in press). An example of a homograph is "nail" which could be interpreted as the nail of one's finger or the nail designed to be hammered.

A source of vocabulary confusion such as homographs would especially affect analogical relations. If terms in the stem are not understood, analogical relationships cannot be established.

Each pair of words in the analogy item (stem, key and each distractor) was classified as either having or not having a homograph. Values for homograph classifications could be either 0 (no homograph) or 1 (homograph). Homographs were identified by the same four people and as with vertical relations, disagreements were resolved by consensus.

Stem-key combinations were also considered for homographs since, when they are located in both, the combination might be a stronger source of confusion. Due to the minimal numbers of items (4 items in 3I and 1 item in 5E) that had this combination, results are not reported. Differential item function medians for homographs in the stem or in the key are presented in Table 8.

Insert Table 8 about here

Results indicate that where there is a homograph in the stem or in the key, Black students tend to do differentially worse (i.e. there is negative DIF). These results are consistent across forms and do not seem to be as interdependent on item position within Verbal section as do other factors.

All homographs in a particular item were added to produce a total homograph value. This value included homographs in the stem, key and distractors. In order to deal with categories that had no values for most forms, when total homograph values were three or more they were collapsed; in such cases the homograph level is represented by a 3. Results of median $DSTD_2$ values for each of these categories across forms and verbal sections are presented in Table 9.

Insert Table 9 about here

Results show that as the number of homographs increases there is a tendency to have more negative DIF. This result is consistent across forms and verbal sections and again does not seem to be as dependent on position as other factors have been.

Semantic Relationships

The general nature of the relationship between the words in the stem of an analogy item can define the type of association that needs to be made in order to correctly identify the option with the same relationship. A number of semantic relationship taxonomies that classify these associations have been developed (Whitely, 1977; Chaffin & Peirce, 1986; Freedle & Gitomer, 1985).

For purposes of this study, since the nature of the relationship was seen as a possible source of differential item functioning, Chaffin and Peirce's (1986) as well as a newer version of Freedle and Gitomer's (1985) semantic relationship taxonomies were used to classify 60 analogy items. Forms OF, 3H, and 4H were classified by Freedle using his taxonomy and by Chaffin and Peirce using theirs. One principal investigator of this study also classified all items of these three forms using both taxonomies. Since the two taxonomies differ more on level of detail than on actual classifications, Chaffin and

Peirce's taxonomy was used for further analyses on DIF. This classification was chosen because it tends to be more inclusive and can be collapsed into five basic relationship types. In addition, two test development specialists and one of the principal investigators classified the previous three forms as well as forms 3I and 5E using Chaffin and Peirce's classification. The discrepancies found were resolved by consensus.

Chaffin and Peirce's taxonomy consists of ten families of relations that were described by Chaffin and Peirce (1986) as follows:

1. CLASS INCLUSION: one word names a class that includes the entity named by the other word. Example: flower:tulip.

2. PART-WHOLE: one word names a part of the entity named by the other word or something that can never be part of the entity named by the other word. Example: tree:forest.

3. SIMILAR: one word represents a different degree or form of the object, action, or quality represented by the other word. Example: breeze:gale.

4. ATTRIBUTE: one word names a characteristic quality, property, or action of the entity named by the other word. Example: glass:fragile.

5. CONTRAST: one word names an opposite or incompatible of the entity named by the other word. Example: default:payment.

6. NON-ATTRIBUTE: one word names a quality, property, or action that is characteristically <u>not</u> an attribute of the entity named by the other word. Example: famine:plentitude.

7. CASE RELATION: one word names an action which the entity named by the other word is usually involved in, or both words name entities that are normally involved in the same action in different ways, e.g., as agent, object, recipient or instrument of the action. Example: doctor:patient.

8. CAUSE/PURPOSE: one word represents the cause, purpose or goal of the entity named by the other word, or the purpose or goal of using the entity named by the other word. Example: joke:laughter.

9. SPACE/TIME: one word names a thing or action that is associated a particular location or time named by the other word. Example: belt:waist.

10. REPRESENTATION: one word names something that is an expression or representation of, or a plan or design for, or provides information about, the entity named by the other word. Example: person:portrait. (pp. 3-5)

For purposes of this study the taxonomy was simplified by collapsing across some of these relations. A five category taxonomy was suggested by Chaffin and it was kept basically the same after it was corroborated as conceptually clear by the test development specialists and principal investigators. The relations collapsed were: 1) Class Inclusion and Part-Whole, 2) Similar and Attribute, 3) Contrast and Non-Attribute, and 4) Case Relation, Cause-Purpose, and Space-Time. The fifth type was considered independent of the other types and was depicted as Representation (the original tenth family relation). Results of the analyses by the simplified taxonomy are presented in Table 10.

Insert Table 10 about here

For each of the five simplified semantic relationship classifications the median of $DSTD_2$ was calculated by form, for V1, V2, and total test. Results show that for some of the five categories the frequency of items in each by form is low. This is especially true for "Representation." The first (Class inclusion & Part-whole), the fourth (Case Relation, Cause-purpose, & Space-time), and the fifth (Representation) classifications have the most consistent negative DIF, while the remaining other classifications, the second (Similar & Attribute) and third (Contrast & Non-attribute), have the most consistent positive DIF. Further inspection shows that again there is an interdependence with position. For the first, fourth, and fifth classifications, the items are mainly located in the first five positions of each analogy section (First: 3I-4/1; 3H-3/0; 4H-2/0; 5E-2/1; Fourth: 3I-5/3; 3H-2/2; 4H-5/5; 5E-6/1, and Fifth: 3I-1/0; 3H-2/0; 4H-1/0; 5E-1/0). For the

second and third classifications, the items identified in these categories are mainly positioned as the last five items of each analogy section (Second: 3I-0/2; 3H-3/4; 4H-2/3; 5E-1/5, and Third: 3I-0/4; 3H-0/4; 4H-0/1; 5E-0/2). Based on these results, no clear conclusions can be reached because of the interdependence between the semantic relationship classifications and position or difficulty.

Distractor Analyses

In order to explore further the previous studied factors, items from each form that had high $DSTD_2$ values have been reproduced in Tables 11 to 14.

Insert Tables 11-14 about here

For each form, items were organized first from the most extreme negatively discrepant, to the least extreme negatively discrepant and second from the most extreme positively discrepant to the least extreme positively discrepant using a $|DSTD_2|$ cutoff of .05. This cutoff was used in order to evaluate a larger number of items. For each of these items, the item number (which indicates position within analogy section, items 36-40 and 16-20 being the first five items of each section) and the actual items with the values of $DSTD_2$ for each distractor as well as for omitted responses are presented. In addition, the actual P value for the Black subgroup, specific characteristics for the test development classifications of subject matter, abstractness level, and the relationship between stem and key, and for the classifications by vertical relationship, homographs and semantic relationships have been identified for each item.

Table 11 presents this information for form 3I. As observed on Figures 1 and 2, only three items have absolute values of DSTD higher than .05 and of

those, two higher than .09. On item 18, the most extreme positive item (DSTD = .20) which is out of bound from the values represented in Figure 2, Black students performed much better. This item includes esoteric vocabulary which is favorable to the Black ethnic background. The word "dashiki" is an African word which represents a type of garment. White students perform differentially worse on this item, and this DIF might be related to less familiarity with the term "dashiki." If White examinees do not know the meaning of this word in the stem, they may not be able to make the analogical connection that is required to answer the item. In addition, the percentage of comparable ability White students who opt to omit this item (17%) is much higher than the comparable percentage of Black students who omit it (5%). Distractor analyses also show that White students were differentially drawn more by the most attractive distractor (B) (17% of Whites versus 13% of Blacks) which has a vertical relationship to the word "garment", the second term of the stem.

Not many items such as item 18, with esoteric vocabulary for White students, are found among SAT analogies, but such an occurrence makes it clear that if vocabulary knowledge peculiar to one subgroup is required to solve analogy items, analogical reasoning skills might not be the main construct tested for that subgroup. Furthermore, the differential attractiveness of distractors which have vertical relationships indicates that, in such items, non-analogical reasoning might come into play.

Item #37 was the only analogy item in form 3I that was differentially more difficult for Black students. Inspection of distractors and item characteristics indicate that Black students tended to differentially omit this

item more than Whites (11% Blacks versus 7% Whites) and that the proponderance of homographs in the stem, key and distractors might have made this item differentially more difficult for Black students. In addition, Black examinees tended to pick the two most attractive distractors B-13% and C-16%. Based on the results, it is interesting to note that on those items that are differentially more difficult for Whites, White students will have a higher tendency to omit while for those items that are differentially more difficult for Blacks, Black students will try to use other strategies and choose a distractor, rather than omitting the item. Prior research on omitting patterns has shown that another minority group, i.e., Hispanics, in general, also omit differentially less than Whites (Rivera & Schmitt, 1986).

Items 39 and 42 are deemed interesting because even though they have virtually no DIF, the distractor analysis shows that different distractors are differentially drawing Black and White students. In item 39, distractor A is differentially more attractive to Blacks while distractor D is more so to Whites. Interestingly, the word "fuel" in distractor A, which is the most attractive distractor in general, has a vertical relationship to the word "petroleum" in the stem, while distractor D has a loosely analogical relationship to the stem (to get an "alloy", you have to "blend"). It would appear that distractors with analogical types of relations might differentially draw more White students while distractors with vertical types of relations might differentially draw more Black students. Item 42 distractor information seems to corroborate this observation. Distractor C with a vertical relationship is differentially more attractive to Black examinees while distractor E, with a very loose associative type of relationship (lots of "trucks" are found on a "highway") is differentially more attractive to White examinees.

Table 12 presents those analogy items in Form 3H which have DIF $\geq |.05|$. As seen in Figures 3 and 4 most DIF seems to occur in the first five items of each analogy set. There are five items which have negative $DSTD_2$ values that are lower than -.05 and two items that have positive $DSTD_2$ values higher than .05. Inspection of these items in Table 12 shows that four of the five negative items are located in the first five positions of either section and that three out of these four items also have Science subject matter content.

Item 16 has the most extreme negative DIF on this form. Inspection of the distractor $DSTD_2$ values shows that distractor C, the most attractive distractor for both groups, differentially draws more Black students. This distractor has a vertical relationship to the stem. In addition, this item also has a homograph in one of the distractors, was categorized as of science content and consists mainly of concrete words. An item similar to this one also appeared as item 19 of Form OF with some variations:

EPIDERMIS:BODY:: (A) stem:plant
(B) air:diver (C) bark:tree
(D) fur:coat (E) nail:finger

The stem of item 16 (Form 3H) is the key of item 19 (Form OF). As it appeared in Form OF, this item also was differentially harder for Black examinees ($DSTD_1$ = -.0775). Distractors D and E were classified as having a vertical relationship on Form OF.

Items 37 and 21, which are differentially harder for Black examinees, have science-related technical terms in the stem. For example, if the term "cumulus" in item 37 is not known, its analogical relationship to "cloud" cannot be made; selection of the correct response would then be more difficult. This same observation can be made for the word "gully" in item 21. Item 21 is also noteworthy in that it is located in the second section set of V2, is a hard item

for Black examinees (P_B = .33) and has been differentially omitted more by this subgroup.

Item 36 distractor analysis provides quite interesting results. Even though this item does not have hard vocabulary, science content or homographs, distractor analysis shows that the two distractors which Black students selected differentially more start with the prefix "auto", which is also part of the first word in the stem. Nevertheless distractor C, which also starts with the prefix "auto", was not selected differentially more by Black examinees. According to this observation, it would appear that, in some cases, non-analogical relations are being used differentially more by Black examinees when answering analogy items.

Item 17 distractor analysis shows that Black examinees chose distractor B differentially. This distractor was identified as having a vertical relationship to the stem. The key and distractor A were also classified as having vertical relationships but also have homographs which might be stronger sources of confusion than the homograph for "play" in distractor B.

The two items in Form 3H that had $DSTD_2$ values higher than .05, indicating that they are differentially easier for Black examinees, were classified as having subject matter content of practical affairs (item 20) and of human relations (item 39). Interestingly, both items are located in one of the first five positions of their respective sections. Apart from neither of these two items including any homographs or having distractors with vertical relationships, there does not seem to be any other explanation for why they are differentially easier for Black examinees.

Corresponding information for Form 4H is presented in Table 13. As observed in Figures 5 and 6, three items have $|DSTD_2|$ values greater than .05, two favoring White and one favoring Black examinees. As observed in Table 13,

in both of the negatively discrepant items (17 and 40) the stem is vertically related to the option which is differentially distracting Black students away from the key. On item 17, 15% of Black students are relating option D - "coach:players" to SOCCER:TEAMS (compared to 12% of the matched White students) rather than perceiving the analogical relationships of "golf:individuals".

On item 40, 21% of Black students (compared to 17% of the matched White students) are attracted to the vertical relationship between ECHO:SOUND and option B "amplifier:speaker". An additional source of confusion here may be the homograph in option B. This item has also been classified as having science content which, as noted earlier, seems to be more difficult for Black examinees.

Interestingly, the one item (#42) on which Black students perform better than expected (despite the vertical relationship between the stem, SYCOPHANT:FLATTERY and option C - "bandit:hypocrisy") is classified as Human Relations. The focal and base groups are also differentially omitting this item (Blacks, 29% and matched Whites, 38%). The large proportion of omitted responses here may be attributable to the difficulty of the vocabulary. Here again White examinees opt to omit differentially.

Item 37 (the second analogy item in V1) is of particular interest because a very similar item is used on Form OF where it appears as item 17 (the second analogy item in V2) as:

STEAM:GEYSER:: (A) power:generator
(B) atoms:reactor (C) coal:mine
(D) lava:volcano (E) rock:quarry

The terms of both the stem and the key are exactly the same but the order has been reversed. The DSTD, value for item 17 in Form OF is -.0383. Recall, however, that only $DSTD_1$ is available for Form OF. The comparable $DSTD_1$ statistic for Form 4H is -.0728. The item is located in the first position (i.e., among the first five items) in 4H-V1. It is in this position that the

effects of differential speededness on analogy items are most extreme. When this is controlled for, the $DSTD_2$ value drops to -.0425 making the items, as they appear in the two forms, of comparable differential difficulty for Black examinees.

Items 39, 43, and 23 are presented as of particular interest due to the observed differential omit patterns. For item 39, responses are characterized by omission for 15% of Blacks and 18% of comparable Whites. The corresponding figures are 16% and 21% for item 43, and 37% and 43% for Item 23.

Figures 7 and 8 indicate five items on Form 5E with $|DSTD_2|$ values greater than .05. Of these, only two are negatively discrepant items. Both of the items with negative $DSTD_2$ values are classified as "Science" (see Table 14), both have vertical relations, and both are in Semantic Relations category 4. On item 17 (the most negatively discrepant), option C, "oar:rowboat", with its vertical relationship to the stem, "CANOE:RAPIDS" is differentially attracting Black examinees. On item 19, option A, "distance:space" is attracting Black students differentially and, again, there is a vertical relationship with the stem, "ACRE:LAND". An additional source of appeal for option A may be the homograph "space".

Of the three items on which Black students did better than expected, items 21 and 41 are classified as having Human Relations subject content and in addition do not have any homographs or vertical relationships. In the third item (20) the key "compile:collection" has a vertical relationship with the stem "CONVENE:ASSEMBLY".

Among the other items of interest listed in Table 14, item 38 was selected because of the differential response patterns of Black and White examinees. Option B, "steel:tin" is vertically related to the stem "ALUMINUM:METAL" and is the most differentially distracting for Black students. On item 39, Whites

differentially select option D, "recital:concert" which has a lose analogical relationship to the stem "REHEARSAL:PLAY", while in option C, "applause:performance" (the response differentially selected by Black students) the term "performance" has a vertical relationship to the term "play" in the stem.

It appears that vertical relationships between the stem and distractors differentially influence the responses of Black examinees, not only on items which have virtually no DIF, but even in those on which Black students do better than expected.

Conclusions

Several factors that seem to be related to the DIF between matched White and Black examinees on 80 SAT analogy items from four test forms have been studied. Results indicate, that after differential speededness is taken into account, the apparent higher negative DIF (i.e., DIF favoring matched White test takers) particular to V1 items is attenuated. On average only one item per form has a DSTD value that is lower than -.10. For each of these extreme negative items, there is a factor or item characteristic that can explain why these items are differentially harder for Black examinees. Vertical relations seem to be the most consistent explanatory characteristic. Items with science content that are positioned among the first five analogy items of either V1 or V2 are also frequently identified as differentially harder for Black examinees but, since there is an interdependence with item position and difficulty, the science effect is harder to separate. Homographs were also identified as a source of confusion that can make vocabulary items harder for subgroups with poor vocabulary skills. These results suggest several points about the performance of comparable Black and White examinees (matched on total SAT-Verbal score).

These are:

- Differential Speededness. Black students seem to take longer to finish tests than White students with comparable total SAT Verbal scores, indicating that speeded tests will be differentially more speeded for Black examinees. This differential speededness might be related to less well developed test-taking skills. Studies of DIF by Ironson and Subkoviak (1979) and by Sennott (1980) showed evidence of differential speededness. This finding does not imply, however, that if given more time students will have better total scores. Research in which testing time was varied for different ethnic groups has shown that the extra time did not differentially help any of the groups (Wild, Durso & Rubin, 1982).
- Position, Item Difficulty, or Subject Content. Items in the first five positions of each analogy section seem to more consistently be identified as differentially harder for Black students. These items are generally the easier items for both groups. The interdependence between subject content ("science" for differentially harder and "human relations" for differentially easier items for Black examinees), item position or difficulty, and level of abstractness makes it difficult to differentiate which factor is most relevant. In addition, the higher proportion of homographs on concrete terms which tend to be part of easier items may make the vocabulary more confusing.
- Vertical Relationships. Distractors that have other than an analogical relationship to the stem, such as a vertical relationship, can differentially attract more Black examinees, particularly when the stem has terms that are difficult or confusing. In cases where vocabulary in the stem is more difficult or confusing for Whites and the item is

differentially more difficult for Whites, White students tend to omit differentially more. Black students, on the other hand, seem to try to guess more and use other strategies (such as vertical relationships) proportionally more than omitting when the item is differentially more difficult for them. When matched White students choose a distractor proportionally more than Black students, they tend to be drawn more toward distractors that have some level of horizontal or analogical association between the terms, rather than toward distractors with vertical relationships. These different strategies by Black and White students in answering analogy items might be related to a differential familiarity with the item type and/or with analogical reasoning by the proportion of the Black population that misses the item.

- o Sources of Vocabulary Confusion. Use of words that might be more esoteric for one subgroup disadvantages their use of analogical reasoning. Correct knowledge of the meaning of a word is imperative in order to make the analogical connections needed to correctly answer analogies. In addition, even for terms that seem relatively easy, sources of vocabulary confusion, such as homographs, might interfere with the understanding of their meaning. Homographs were observed more frequently on those items that were identified as differentially harder for Black examinees.
- o Type of Associative Relationship. Items differentially more difficult for Black examinees tended to test "Case Relations," "Cause-Purpose," or "Space-Time" relationships. However, it would appear that such semantic classifications might not be particularly relevant to the study of DIF for Black examinees, especially if DIF is largely a function of non-analogical reasoning by Black students.

These conclusions should be viewed as tentative. Due to the interdependence of several of the identified factors, and in some cases their limited occurrence, it is imperative that these factors be studied more systematically. A study in which factors such as position, subject content, level of abstractness, vertical relationships, homographs, vocabulary difficulty, and semantic relationships are controlled has been proposed. Once the identified factors are confirmed in a rigorously designed study, new test development guidelines can be developed. The current results, nevertheless, extend our understanding and identify factors that may cause differential item functioning in analogy items.

as\rr96950

References

Bleistein, C. A., & Wright, D. (1986, April). Assessment of unexpected differential item difficulty for Asian-American candidates on the Scholastic Aptitude Test. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Chaffin, R., & Peirce, L. (1986-draft). A taxonomy of Semantic Relations for the classification of GRE Analogy items. Educational Testing Service, Princeton, N.J.

Dorans, N. J. (1982). Technical review of item fairness studies: 1975-1979. (Research Report No. 82-90). Educational Testing Service, Princeton, NJ.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. Journal of Educational Measurement.

Dorans, N. J., & Kulick, E. (1983). Assessing unexpected differential item performance of female candidates on SAT and TSWE forms administered in December 1977: An application of the standardization approach. (Research Report No. RR-83-9). Princeton, NJ: Educational Testing Service.

Dorans, N. J., & Schmitt, A. P. (1986-draft). Issues in DIF analyses for the SAT: Differential Speededness and the Choice of Proportion Correct. Princeton, NJ.

Echternacht, G. (1972, March). An examination of test bias and response characteristics for six candidate groups taking the ATGSB (PR-72-4). Princeton, NJ: Educational Testing Service. (ED 065 510)

Freedle, R., & Gittomer, D. (1985, October). Interim report for PRPC project 969-43. Internal ETS memorandum for A. Schmitt and N. Dorans.

Ironson, G. H., & Subkoviak, M. J. (1979). A comparison of several methods of assessing item bias. Journal of Educational Measurement, 16, 209-225.

Kulick, E. (1984, July). Assessing unexpected differential item performance of Black candidates of SAT Form CSA6 and TSWE Form E33 (SR-84-80). Princeton, NJ: Educational Testing Service.

Rivera, C., & Schmitt. A. (1986, April). A comparison of Hispanic and White students' omit patterns on the Scholastic Aptitude Test. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Rogers, H. J., & Kulick, E. (1986, April). An investigation of unexpected differences in item performance between Blacks and Whites taking the SAT. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Scheuneman, J. D. (1978). Ethnic group bias in intelligence test items. In S. W. Lundsteen (Ed.), Cultural factors in learning and instruction. New York: ERIC Clearinghouse on Urban Education, Diversity Series, No. 56.

Schmitt, A. P. (in press). Unexpected differential item performance of Hispanic examinees. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Shepard, L. A., Camilli, G., & Williams, D. M. (1984). Accounting for statistical artifacts in item bias research. Journal of Educational Statistics, 9, 93-128.

Sinnott, L. T. (1980, August). Differences in item performance across groups (RR-80-19). Princeton, NJ: Educational Testing Service.

Stricker, L. J. (1982). Identifying test items that perform differently in population subgroups: A partial correlation index. Applied Psychological Measurement, 6, 261-273.

Whitely, S. E. (1977). Relationships in analogy items: A semantic component of a psychometric task. Educational and Psychological Measurement, 37, 725-739.

Wild, C. L., Durso, R., & Rubin, D. B. (1982). Effect of increased test-taking time on test scores by ethnic group, years out of school, and sex. Journal of Educational Measurement, 19, 19-28.

Table 1

Summary of Sampling, Scaled Score Means, and Reliability Information

Form	Sample Size		Scaled Score Mean	Reliability
	Black	White		
OF	5,971[1]	65,895[1]	431	.921
3H	21,789[2]	278,099[2]	436	.926
4H	23,115[2]	285,883[2]	440	.922
3I	8,757[1]	25,589[1]	399	.921
5E	5,993[1]	35,218[1]	436	.917

Note: All five SAT forms are disclosed forms.

[1]Representative Samples from Total Population

[2]Total Black and White Populations for whom English is the best language

Table 2

Median Values of DIF Indices for Analogy Items
by Verbal Section, Item Position, and Forms

Item Position	Forms							
	3I		3H		4H		5E	
	$DSTD_1$	$DSTD_2$	$DSTD_1$	$DSTD_2$	$DSTD_1$	$DSTD_2$	$DSTD_1$	$DSTD_2$
Verbal 1 (n=10)								
1-5	-.04	-.02	-.07	-.05	-.07	-.04	-.03	-.02
6-10	-.00	.01	.01	.03	-.01	-.00	.01	.02
All	-.02	-.01	.00	.02	-.02	-.01	-.00	.01
Verbal 2 (n=10)								
1-5	-.00	-.00	-.00	-.00	-.03	-.03	-.04	-.04
6-10	.01	.01	-.00	-.00	.02	.02	.02	.02
All	.00	.00	-.00	-.00	.01	.01	.00	.00
Total Test (n=20)								
1-5	-.02	-.01	-.05	-.04	-.03	-.03	-.03	-.03
6-10	.00	.01	.01	.02	.02	.02	.01	.02
All	-.01	-.00	-.00	.00	-.01	-.00	-.00	.01

Note. Position 1-5 refers to items 36-40 in Verbal 1 and items 16-20 in Verbal 2. Position 6-10 refers to items 41-45 in Verbal 1 and items 21-25 in Verbal 2.

Note. $DSTD_1$ is the standardized p-difference based on all Black and White students for whom English was their best language who took the examinations. $DSTD_2$ calculations exclude students who did not reach the item.

Table 3

Summary of Items Flagged as Differentially Difficult for Black Examinees by Form, Item Type, and DSTD

Form	Item Type	$DSTD_1$ -.10<DSTD≤-.05 F (%)	$DSTD_1$ ≤-.10 F (%)	$DSTD_2$ -.10<DSTD≤-.05 F (%)	$DSTD_2$ ≤-.10 F (%)
3I	Antonyms (n=25)	2(8)	1(4)	2(8)	1(4)
	Analogy (n=20)	1(5)	1(5)	0(0)	1(5)
	Sentence Comp. (n-15)	0(0)	0(0)	0(0)	0(0)
	Reading Comp. (n-25)	1(4)	0(0)	0(0)	0(0)
3H	Antonym (n=25)	1(4)	0(0)	1(4)	0(0)
	Analogy (n=20)	5(25)	1(5)	4(20)	1(5)
	Sentence Comp. (n=15)	0(0)	0(0)	0(0)	0(0)
	Reading Comp. (n=25)	0(0)	0(0)	0(0)	0(0)
4H	Antonym (n=25)	1(4)	1(4)	1(4)	1(4)
	Analogy (n=20)	4(20)	0(0)	2(10)	0(0)
	Sentence Comp. (n=15)	0(0)	0(0)	0(0)	0(0)
	Reading Comp. (n=25)	1(4)	0(0)	1(4)	0(0)
5E	Antonym (n=25)	2(8)	0(0)	2(8)	0(0)
	Analogy (n=20)	3(15)	0(0)	2(10)	0(0)
	Sentence Comp. (n=15)	2(13)	0(0)	2(13)	0(0)
	Reading Comp. (n=25)	2(8)	0(0)	2(8)	0(0)

Note: $DSTD_1$ is the standardized p-difference based on all Black and White students for whom English was their best language who took the examinations. $DSTD_2$ calculations exclude students who did not reach the item.

Table 4

Median Values of $DSTD_2$ for Analogy Items
by Section and Form for Subject Matter Classifications

	Form							
	3I		3H		4H		5E	
Section	n	Median	n	Median	n	Median	n	Median
Aesthetic/Philosophical								
V1	4	-.01	3	.01	2	.01	2	.02
V2	2	.01	3	-.00	3	.02	3	.02
Total	6	.00	6	.01	5	.02	5	.02
World of Practical Affairs								
V1	2	.01	1	.03	3	-.00	3	-.02
V2	3	.01	3	.01	2	-.04	2	.02
Total	5	.01	4	.02	5	-.00	5	-.01
Science								
V1	2	-.03	3	-.05	3	-.04	2	-.03
V2	3	-.02	2	-.08	3	.01	3	-.07
Total	5	-.02	5	-.07	6	-.02	5	-.03
Human Relationships								
V1	2	.02	3	.02	2	.01	3	.03
V2	2	.01	2	.02	2	-.00	2	.04
Total	4	.01	5	.02	4	-.00	5	.03

Table 5

Median Values of $DSTD_2$ for Analogy Items by Section, Form, and Word Abstractness

	Form							
	3I		3H		4H		5E	
Section	n	Median	n	Median	n	Median	n	Median
Concrete								
V1	2	-.02	2	-.06	3	-.04	2	-.02
V2	4	-.00	3	-.06	2	-.02	3	-.04
Total	6	-.00	5	-.06	5	-.03	5	-.02
Mixed								
V1	4	-.02	4	.02	5	-.00	4	-.01
V2	3	-.02	3	.01	4	.02	4	.05
Total	7	-.02	7	.01	9	.01	8	.01
Abstract								
V1	4	.01	4	.03	2	.01	4	.03
V2	3	.02	4	.02	4	.01	3	.01
Total	7	.01	8	.03	6	.01	7	.02

Table 6

Median Values of $DSTD_2$ for Analogy Items
by Section, Form, and Relationship Between Stem and Key

	Form							
	3I		3H		4H		5E	
Section	n	Median	n	Median	n	Median	n	Median
Independent								
V1	6	-.01	8	.02	6	-.01	5	.02
V2	7	.01	8	.00	7	.02	7	.00
Total	13	-.00	16	.02	13	-.00	12	.00
Overlapping								
V1	4	-.00	2	-.02	4	-.02	5	.01
V2	3	-.01	2	-.08	3	-.01	3	.06
Total	7	-.01	4	-.06	7	-.01	8	.01

Table 7

Median Values of $DSTD_2$ for Analogy Items
by Section, Form, and Vertical Relationships

Section	Vert. Code	Form 3I n	3I Median	3H n	3H Median	4H n	4H Median	5E n	5E Median
				Key					
V1	0	6	-.01	7	.03	6	-.02	6	.02
	1	4	-.00	3	-.03	4	.00	4	-.00
V2	0	5	-.01	7	-.00	7	.02	8	.00
	1	5	.01	3	-.00	3	-.03	2	.02
Total	0	11	-.01	14	.01	13	-.00	14	.01
	1	9	.01	6	-.02	7	-.00	6	-.00
				Distractors					
V1	0	4	.01	4	.03	4	-.00	5	.01
	1	5	-.01	6	-.01	4	-.00	1	.03
	2	1	-.05	0	-	1	-.04	4	-.03
	3	0	-	0	-	1	-.03	0	-
V2	0	5	-.01	2	.05	2	.01	4	.04
	1	5	.01	2	-.00	4	.02	3	-.04
	2	0	-	6	-.04	3	.01	3	-.01
	3	0	-	0	-	1	-.03	0	-
Total	0	9	-.00	6	.04	6	-.00	9	.02
	1	10	.00	8	-.00	8	.01	4	-.01
	2	1	-.05	6	-.04	4	-.02	7	-.02
	3	0	-	0	-	2	-.03	0	-

Table 8

Median Values of $DSTD_2$ for Analogy Items
by Section, Form, and Homographs in Stem and Key

Section	Hom. Code	Form 3I n	3I Median	3H n	3H Median	4H n	4H Median	5E n	5E Median
				Stem					
V1	0	6	-.01	10	.02	7	-.02	7	.01
	1	4	-.00	0	-	3	-.00	3	.01
V2	0	9	.01	8	-.00	9	.02	9	.01
	1	1	-.02	2	-.04	1	-.03	1	-.04
Total	0	15	-.00	18	.01	16	.00	16	.01
	1	5	-.01	2	-.04	4	-.02	4	-.02
				Key					
V1	0	5	-.00	9	.02	9	-.00	8	.01
	1	5	-.01	1	-.03	1	-.06	2	-.00
V2	0	10	.00	8	.00	10	.01	8	-.00
	1	0	-	2	-.06	0	-	2	.01
Total	0	15	-.00	17	.01	19	-.00	16	.01
	1	5	-.01	3	-.06	1	-.06	4	.01

Table 9

Median Values of $DSTD_2$ for Analogy Items
by Section, Form, and Total Homographs

Section	Hom. Code	Form 3I n	3I Median	3H n	3H Median	4H n	4H Median	5E n	5E Median
V1	0	4	.01	9	.02	4	-.01	4	.00
	1	1	-.00	1	-.03	2	-.01	4	.01
	2	4	-.00	0	-	3	.01	1	-.02
	3	1	-.13	0	-	1	-.06	1	.01
V2	0	7	-.00	3	.05	8	.02	4	.01
	1	1	.01	4	-.01	2	-.03	5	.01
	2	1	.03	2	-.05	0	-	1	-.07
	3	1	-.02	1	-.06	0	-	0	-
Total	0	11	-.00	12	.03	12	.01	8	.01
	1	2	.01	5	-.02	4	-.03	9	.01
	2	5	.01	2	-.05	3	.01	2	-.05
	3	2	-.08	1	-.06	1	-.06	1	.01

Table 10

Median Values of $DSTD_2$ for Analogy Items
by Section, Form, and Semantic Relationship Classification

	Form							
	3I		3H		4H		5E	
Section	n	Median	n	Median	n	Median	n	Median
Class Inclusion & Part-Whole								
V1	2	-.01	1	-.08	2	.01	2	-.03
V2	3	-.01	2	-.08	0	-	1	.04
Total	5	-.01	3	-.08	2	.01	3	-.02
Similar & Attribute								
V1	2	.01	4	-.01	2	.03	1	.01
V2	0	-	3	.01	3	.01	5	.02
Total	2	.01	7	.01	5	.01	6	.02
Contrast & Non-Attribute								
V1	0	-	2	.03	1	-.02	2	.03
V2	4	.01	2	.01	1	.02	0	-
Total	4	.01	4	.03	2	-.00	2	.03
Case Relations, Cause-Purpose & Space-Time								
V1	6	-.00	2	.03	5	-.04	3	.01
V2	2	-.01	2	-.03	5	.02	4	-.06
Total	8	-.01	4	.01	10	-.02	7	-.04
Representation								
V1	0	-	1	-.06	0	-	2	.02
V2	1	-.00	1	.06	1	-.03	0	-
Total	1	-.00	2	-.00	1	-.03	2	.02

Table 11

SAT Analogy Items from Form 3I
with High Differential Item Functioning

Item No.	Item	$DSTD_2$	P_B[1]	Characteristics: TD Class.	Vertical Rel.	Homograph	Semantic[2] Rel.
37	WHITTLE:STICK:: .02 (A) lick:hand .03 (B) shatter:glass .02 (C) hammer:nail -.13 (D) chisel:stone .03 (E) nuzzle:nose .04 Omitted	-.1316	.49	Aesthetic Mixed Independent	Key	Stem Key A C	4
18	DASHIKI:GARMENT:: .20 (A) spoon:utensil -.04 (B) hat:coat -.01 (C) cotton:summer .00 (D) foot:shoe -.02 (E) plate:table -.12 Omitted	.2006	.70	Practical Affairs Concrete Independent	B	0	1
41	TIDINGS:MESSENGER:: -.04 (A) clue:detective .06 (B) gossip:telltale .00 (C) verdict:convict .00 (D) inauguration:voter .00 (E) exposure:spy -.01 Omitted	.0563	.38	Human Mixed Overlapping	0	0	4
	Other Items of Interest						
39	REFINE:PETROLEUM:: .03 (A) consume:fuel -.02 (B) smelt:ore .00 (C) prospect:uranium -.03 (D) blend:alloy .01 (E) import:rubber .02 Omitted	-.0199	.40	Science Mixed Overlapping	Key A	0	4
42	COMPLEX:BUILDING:: -.01 (A) tapestry:fabric .02 (B) apple:tree .03 (C) classroom:campus -.01 (D) federation:state -.04 (E) highway:truck .01 Omitted	-.0138	.26	Aesthetic Mixed Independent	C	Stem Key	1

[1] P_B is the actual proportion correct for the Black subgroup.

[2] Values preceding options represent $DSTD_2$ based on distractor analysis. The numbers under "Semantic Rel." represent the following classifications: 1 - Class inclusion and Part-whole, 2 - Similar and Attribute, 3 - Contrast and Non-attribute, 4 - Case relation, Cause-purpose, and Space-time, and 5 - Representation.

Table 12

SAT Analogy Items from Form 3H
with High Differential Item Functioning

Item No.	Item	$DSTD_2$	P_B[1]	Characteristics: TD Class.	Vertical Rel.	Homograph	Semantic[2] Rel.
16	BARK:TREE:: -.10 (A) skin:fruit .01 (B) dew:grass .06 (C) seed:flower .03 (D) peak:hill .01 (E) wake:boat .00 Omitted	-.1009	.74	Science Concrete Overlapping	C	Stem E	1
37	CUMULUS:CLOUD:: .01 (A) lake:ocean .01 (B) carnivore:meat .02 (C) glacier:blizzard -.08 (D) evergreen:tree .02 (E) evening:daylight .02 Omitted	-.0801	.48	Science Mixed Independent	0	0	1
21	GULLY:EROSION:: .01 (A) drought:precipitation -.07 (B) mine:excavation -.00 (C) clot:dispersion -.01 (D) forest:cultivation .00 (E) water:inundation .07 Omitted	-.0653	.33	Science Mixed Independent	A E	Key	4
36	AUTOBIOGRAPHY:AUTHOR:: .02 (A) autograph:signature .01 (B) self-sufficiency: provision .00 (C) automation:worker -.06 (D) self-portrait:artist .03 (E) autopsy:doctor .00 Omitted	-.0624	.63	Aesthetic Concrete Overlapping	Key A	0	5
17	EXCERPT:BOOK:: .01 (A) type:page .04 (B) script:play .00 (C) solo:routine -.06 (D) clip:film .00 (E) drama:musical .01 Omitted	-.0563	.60	Aesthetic Concrete Overlapping	Key A B	Key A B	1
20	BLUEPRINT:BUILDING:: -.02 (A) receipt:money -.02 (B) symphony:concert -.02 (C) map:automobile -.01 (D) briefcase:lawyer .06 (E) agenda:meeting .01 Omitted	.0612	.65	Practical Affairs Mixed Independent	0	0	5
39	SOLITUDE:RECLUSE:: .05 (A) attention:exhibitionist -.02 (B) courtesy:braggart .00 (C) poverty:donor .00 (D) abuse:official .00 (E) persecution:director -.04 Omitted	.0511	.38	Human Rel. Mixed Independent	0	0	4

[1] P_B is the actual proportion correct for the Black subgroup.

[2] Values preceding options represent $DSTD_2$ based on distractor analysis. The numbers under "Semantic Rel." represent the following classifications: 1 - Class inclusion and Part-whole, 2 - Similar and Attribute, 3 - Contrast and Non-attribute, 4 - Case relation, Cause-purpose, and Space-time, and 5 - Representation.

Table 13

SAT Analogy Items from Form 4H
with High Differential Item Functioning

Item No.	Item	$DSTD_2$	P_B^1	Characteristics: TD Class.	Vertical Rel.	Homograph	Semantic[2] Rel.
17	SOCCER:TEAMS:: -.09 (A) golf:individuals .01 (B) football:spectators .02 (C) badminton:rackets .03 (D) coach:players .02 (E) baseball:bases .01 Omitted	-.0863	.67	Practical Affairs Mixed Overlapping	B D	0	4
40	ECHO:SOUND:: .00 (A) outline:image .04 (B) amplifier:speaker -.06 (C) reflection:light .00 (D) deflection:blow .00 (E) resistor:electricity .01 Omitted	-.0592	.48	Science Mixed Overlapping	Key B	Key B D	4
42	SYCOPHANT:FLATTERY:: .00 (A) imposter:deference .00 (B) embezzler:insolence .01 (C) bandit:hypocrisy .07 (D) swindler:fraudulence .00 (E) advocate:defamation -.09 Omitted	.0659	.23	Human Rel. Mixed Independent	C	0	2
	Other Items of Interest						
37	GEYSER:STEAM:: .00 (A) fountain:coin -.04 (B) volcano:lava .02 (C) glacier:iceberg .01 (D) avalanche:trees .00 (E) mudslide:rocks .01 Omitted	-.0425	.65	Science Concrete Independent	0	0	4
39	PARAGRAPH:PROSE:: .03 (A) stanza:poetry .00 (B) tenor:harmony -.01 (C) refrain:chorus .00 (D) signature:letter .02 (E) edition:book -.03 Omitted	.0259	.38	Aesthetic Abstract Overlapping	Key	C D	1
43	ABYSS:DITCH:: .00 (A) conflagration:campfire .02 (B) velocity:acceleration .01 (C) umbrella:rain .01 (D) square:rectangle .01 (E) basement:house -.05 Omitted	-.0036	.12	Science Concrete Independent	0	Stem	2
23	PURITANICAL:FRIVOLITY:: .00 (A) secretive:ascendancy .01 (B) neurotic:phobia .01 (C) intrepid:courage .02 (D) religious:doctrine .02 (E) lazy:exertion -.06 Omitted	.0158	.21	Aesthetic Abstract Independent	D	0	3

[1] P_B is the actual proportion correct for the Black subgroup.

[2] Values preceding options represent $DSTD_2$ based on distractor analysis. The numbers under "Semantic Rel." represent the following classifications: 1 - Class inclusion and Part-whole, 2 - Similar and Attribute, 3 - Contrast and Non-attribute, 4 - Case relation, Cause-purpose, and Space-time, and 5 - Representation.

Table 14

SAT Analogy Items from Form 5E
with High Differential Item Functioning

Item No.	Item	$DSTD_2$	P_B[1]	Characteristics: TD Class.	Vertical Rel.	Homograph	Semantic Rel.[2]
17	CANOE:RAPIDS:: -.08 (A) plane:turbulence .00 (B) truck:garage .05 (C) oar:rowboat .01 (D) factory:automation .01 (E) pond:stream .02 Omitted	-.0825	.66	Science Concrete Independent	C	0	4
19	ACRE:LAND:: .05 (A) distance:space .01 (B) speed:movement -.07 (C) gallon:liquid .01 (D) degree:thermometer .00 (E) year:birthday .01 Omitted	-.0746	.46	Science Mixed Independent	A	A D	4
21	COMPATRIOTS:COUNTRY:: .00 (A) transients:home .06 (B) kinsfolk:family -.06 (C) competitors:team .00 (D) performers:audience -.01 (E) figureheads:government .01 Omitted	.0640	.32	Human Rel. Mixed Overlapping	0	0	2
20	CONVENE:ASSEMBLY:: .01 (A) borrow:library .00 (B) reprove:defiance .00 (C) contrast:shadow -.02 (D) implicate:court .06 (E) compile:collection -.04 Omitted	.0601	.46	Practical Affairs Mixed Overlapping	Key	D	2
41	APOLOGY:RUEFUL:: -.02 (A) confession:inquisitive -.04 (B) request:grateful .01 (C) recommendation:censorious .06 (D) boast:proud .01 (E) taunt:timid -.02 Omitted	.0587	.26	Human Rel. Abstract Independent	0	0	5
Other Items of Interest							
38	ALUMINUM:METAL:: -.01 (A) alloy:element .04 (B) steel:tin .00 (C) hydrogen:water -.03 (D) oxygen:gas .00 (E) diamond:ring .00 Omitted	-.0338	.61	Science Mixed Independent	A B	0	1
39	REHEARSAL:PLAY:: .01 (A) draft:essay .01 (B) manual:process .04 (C) applause:performance -.06 (D) recital:concert .01 (E) journal:news .00 Omitted	.0123	.59	Aesthetic Mixed Overlapping	C D	Stem Key B	4

[1] P_B is the actual proportion correct for the Black subgroup.

[2] Values preceding options represent $DSTD_2$ based on distractor analysis. The numbers under "Semantic Rel." represent the following classifications: 1 - Class inclusion and Part-whole, 2 - Similar and Attribute, 3 - Contrast and Non-attribute, 4 - Case relation, Cause-purpose, and Space-time, and 5 - Representation.

FIGURE 1 – FORM 3I / V1 ANALOGY ITEMS
STANDARDIZED DIFFERENTIAL ITEM FUNCTION VALUES

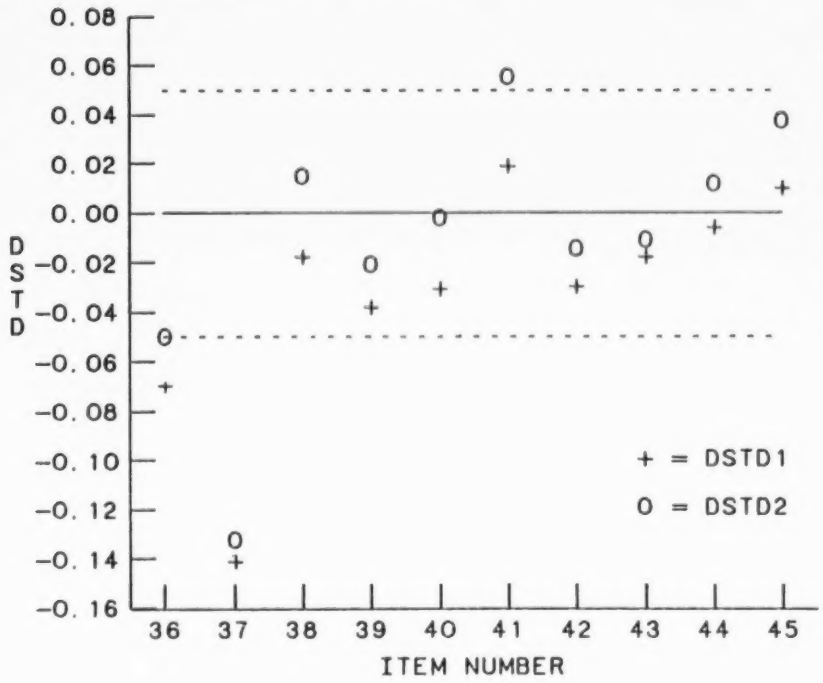
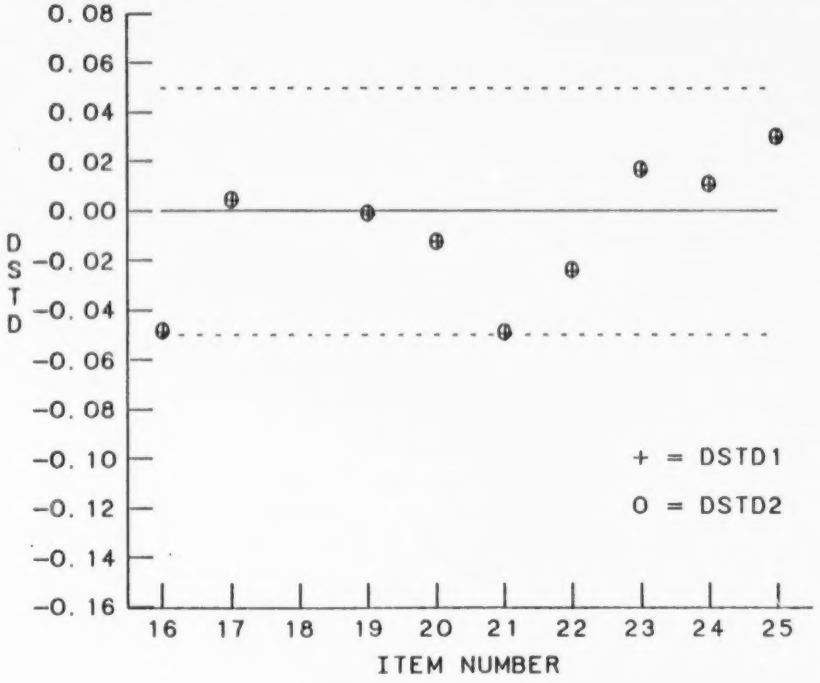


FIGURE 2[1] – FORM 3I / V2 ANALOGY ITEMS
STANDARDIZED DIFFERENTIAL ITEM FUNCTION VALUES



[1]DSTD value for item 18 on V2 is .20 and is out of bound from the values represented on Figure 2.

FIGURE 3 - FORM 3H / V1 ANALOGY ITEMS
STANDARDIZED DIFFERENTIAL ITEM FUNCTION VALUES

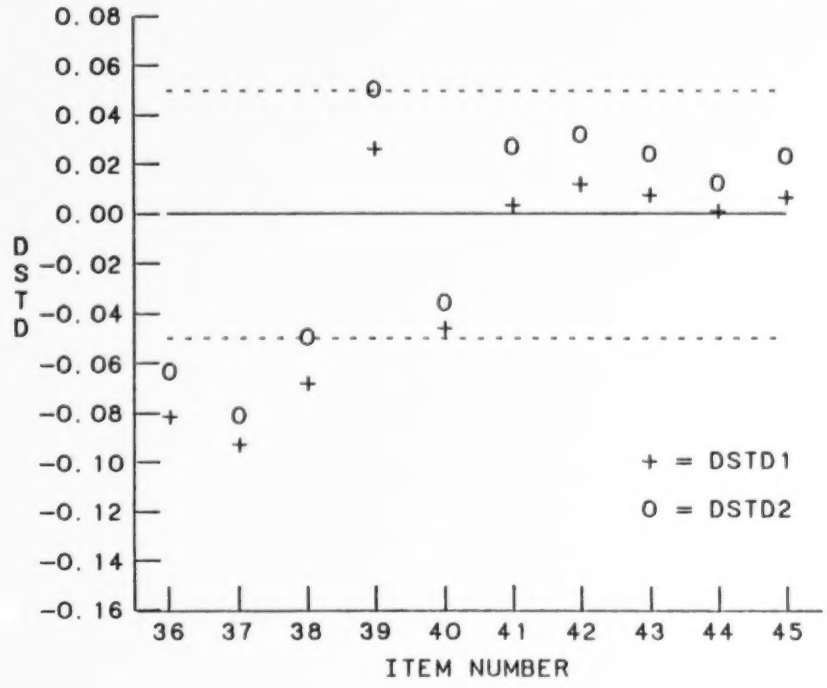


FIGURE 4 - FORM 3H / V2 ANALOGY ITEMS
STANDARDIZED DIFFERENTIAL ITEM FUNCTION VALUES

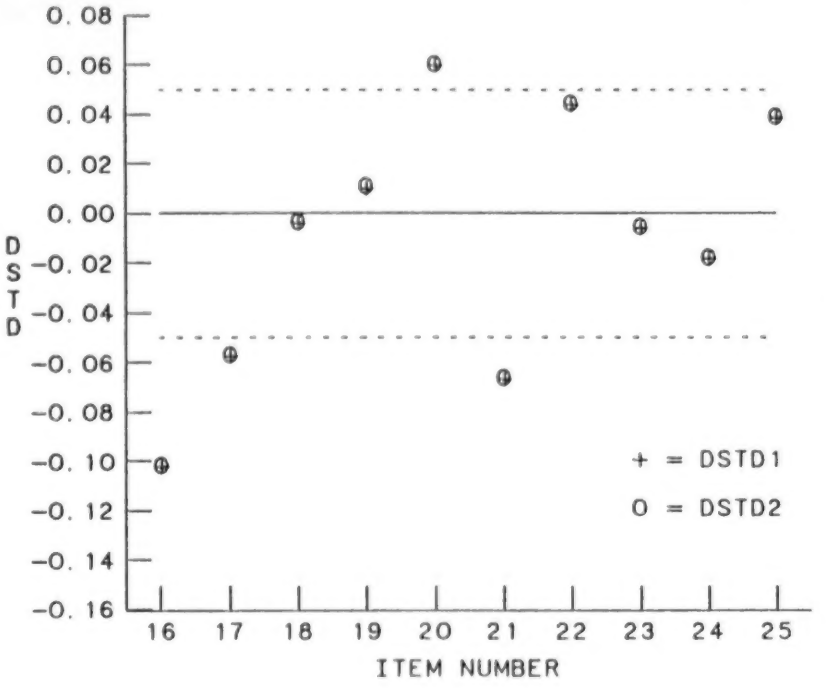
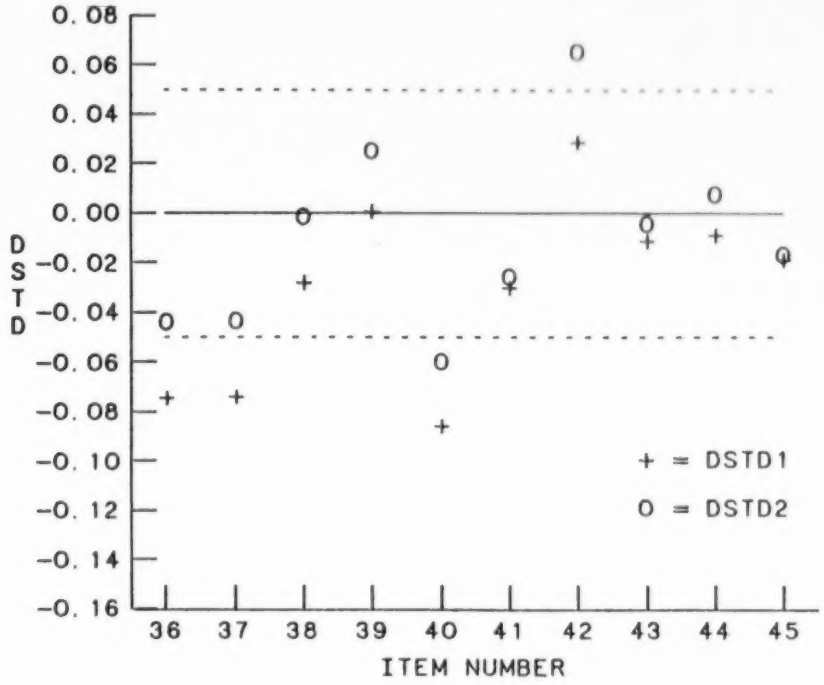
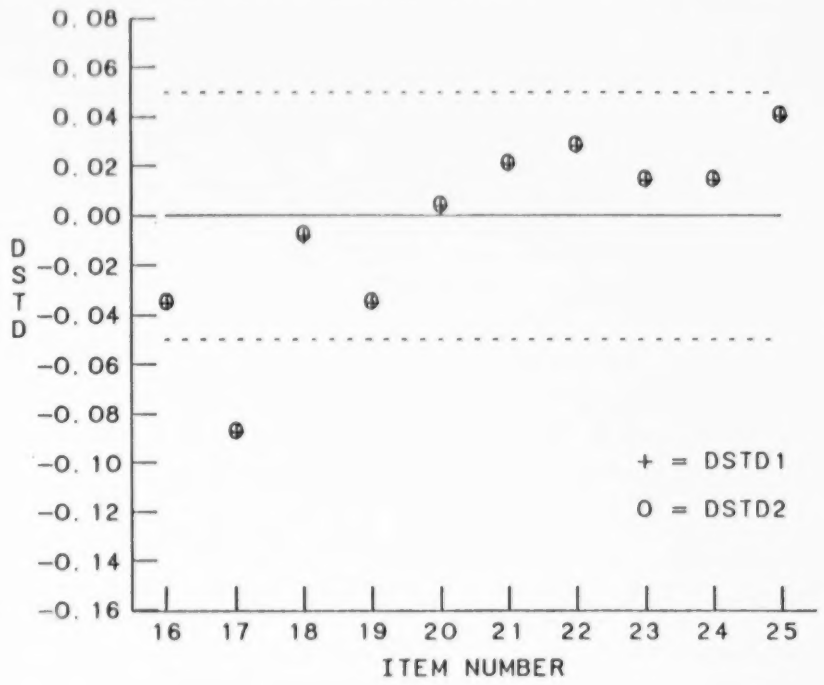


FIGURE 5 - FORM 4H / V1 ANALOGY ITEMS
STANDARDIZED DIFFERENTIAL ITEM FUNCTION VALUES



53

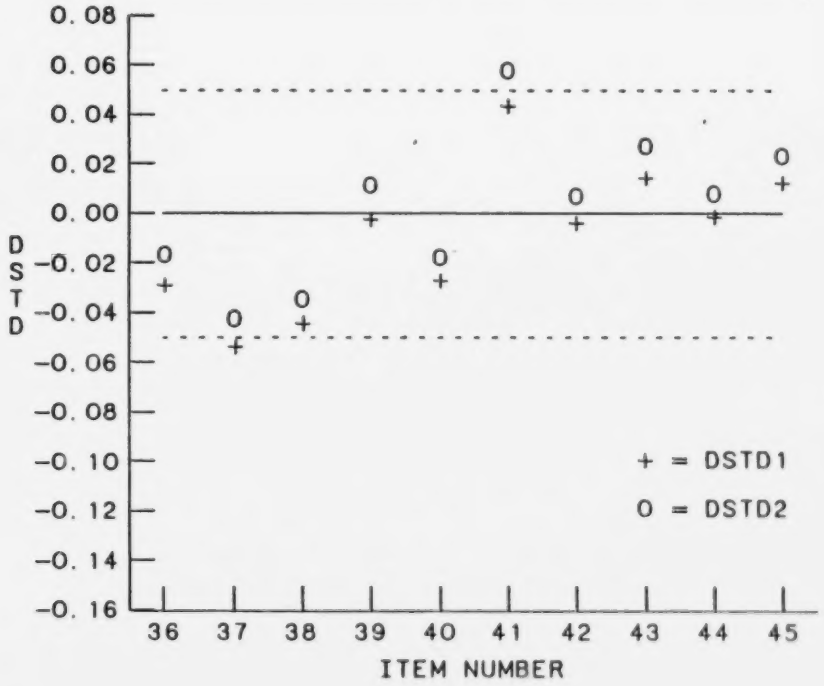
FIGURE 6 - FORM 4H / V2 ANALOGY ITEMS
STANDARDIZED DIFFERENTIAL ITEM FUNCTION VALUES



45

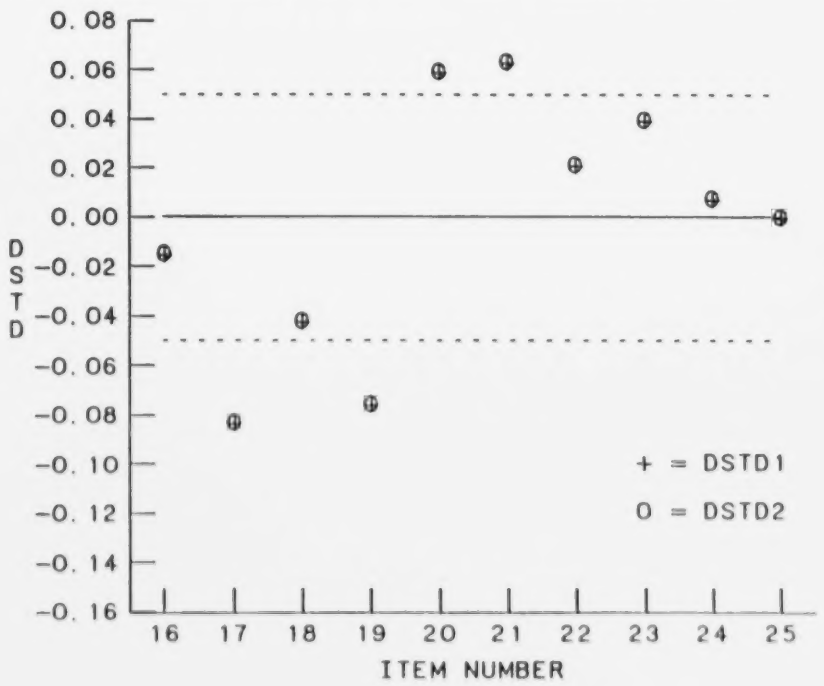
54

FIGURE 7 - FORM 5E / V1 ANALOGY ITEMS
STANDARDIZED DIFFERENTIAL ITEM FUNCTION VALUES



55

FIGURE 8 - FORM 5E / V2 ANALOGY ITEMS
STANDARDIZED DIFFERENTIAL ITEM FUNCTION VALUES



46

56